BIBLID 1133-1127 (2011) p. 253-294

Received 21 June 2011 · Accepted 19 July 2011

A diachronic approach to scientific lexicon in English: Evidence from Late Modern English corpora*

Pascual Cantos and Nila Vázquez · Universidad de Murcia (Spain)

ABSTRACT

This research focuses on the *Corpus of English Texts on Astronomy* (CETA), the first sub-corpus of CC, presenting a diachronic approach to astronomy specific lexicon found in texts from 1710 to 1920. The goal of this research is trying to determine the evolution of the lexical astronomy-domain specificity in the CETA. That is, how many astronomy-like lexical features occur in the English astronomic texts gathered in the CETA. This might shed some light on: (i) the introduction rate of new astronomic specific vocabulary along time, (ii) lexical richness in English astronomic texts, (iii) the rate of new astronomic specific vocabulary along time, (iv) the potential lexical specific features of English astronomic texts, and (v) lexico-semantic text difficulty of English astronomic texts.

Keywords: Diachronic corpus-based research, corpus linguistics, astronomy, lexical complexity, lexical specificity.

---

* We gratefully acknowledge funding from the following institutions: Fundación Séneca (grant 08594/PHCS/08Ñ) and Spanish Ministry for Science and Innovation (grant HUM2007-60706).

1. Introduction

The Late Modern English period has been observed by some scholars (Rydén, 1984; Denison 1998) to be the most neglected period in the history of the English language. However, this contrasts with the fact that this period is a very well-documented one, and is much more easily accessible to the speaker of Present-day English than –say– the Old or Middle English periods. In addition, new corpora have been compiled to fill the gap between Early Modern English and Present-day English. The *Lampeter Corpus*, for instance, deals with the shift from Early to Late Modern English; the *Corpus of Late Modern English Prose* covers the latter half of the 19th and the beginning of the 20th centuries; the *ARCHER Corpus* comprises the entire period from Late Modern to Present-day English and some more could be mentioned here.[1]

The *Coruña Corpus* (CC) is a good tool to analyze specific purpose language in early periods of English: It consists of a collection of samples for the historical study of English scientific writing on which the MUSTE[2] Research Group has been working since 2003 in the University of A Coruña (Spain). The CC aims at becoming a reference for the study of linguistic change and variation

---

1 Further information on Historical English Corpora can be found in Vázquez *et al.* (forthcoming).

2 Research Group for Multidimensional Corpus-Based Studies in English (http://www.udc.es/grupos/muste/index.html).

in English scientific writing in general and among different scientific domains and disciplines. Chronologically, texts range from 1700 to 1900 and it offers a wonderful opportunity to study the scientific register and style from a diachronic and synchronic point of view.[3]

The present research will focus on the *Corpus of English Texts on Astronomy* (CETA), the first sub-corpus of CC, presenting a diachronic approach to astronomy specific lexicon found in texts from 1710 to 1920.

Astronomy is probably one of the oldest sciences, interested in the study of celestial objects and phenomena that originate outside the Earth's atmosphere. It is concerned with the evolution, physics, chemistry, meteorology, and motion of celestial objects, as well as the formation and development of the universe. Astronomers have always performed methodical observations of the night sky, and astronomical artefacts have been found from much earlier periods.

2. Domain specific lexical features

Variation in language is conditioned by "uses" rather than by "users" (Romaine, 1994, p. 20; Moskowich & Crespo, 2009, p. 47) and it is precisely the need for new uses what makes languages change and adopt specific vocabulary to refer to innovative items. Some fields demand a particular type of lexicon. Whenever it is absent from a language, a process of borrowing must take place. In the case of English, Latin (itself or through French) becomes its main model and the main source for scientific lexicon. The vernacularization of medical writing in the fourteenth century sets the starting point for some sort of 'English for Specific Purposes' which will extend to other scientific subfields in the course of time.[4]

---

[3] For a more detailed description of the *Coruña Corpus* see Moskowich-Spiegel Fandiño & Crespo 2007: 341-357 and Moskowich-Spiegel Fandiño, I. (forthcoming).

[4] Diachronic studies dealing with scientific vocabulary are scarce, though we can mention Atkinson (1992, 1996), Norri (1992, 1998) and Taavitsainen (2001, 2002) or, more recently in Spain, Lareo-Martín & Montoya-Reyes (2007), Alonso-Almeida & Sánchez-Cuervo (2009) and Moskowich & Crespo (2009), among others.

Terminology is commonly claimed to be the study of terms and their use. Within this discipline, terms are words and compound words that are used in specific contexts. Terminology is located at the crossroads of a large number of subdisciplines of linguistics (Cabré, 1998, p. 11). It is the shortened form of 'technical terms' or 'terms of art', which are identified within a discipline or speciality field. Cabré (1993, p. 170) defines terms as form and content units which belong to the system of a certain language, having various alternative specific subsystems coexisting in their interior. Therefore, terminology systematically studies the 'designation of concepts' particular to one or more domains of human activity, through research and analysis of terms in context, for the purpose of documenting and promoting correct usage. In this way, a technical term could be defined as a textual realization of a specialized concept (Spasic *et al.,* 2005, p. 240). Within special subject languages (SL), general language (GL) words or non-terms coexist with these technical terms, being the latter lacking in idiosyncratic linguistic features which formally distinguish them from the former. In this respect, Sager (1990, p. 9) claims that some terms are exclusive to a particular domain, such as the zoological term "orbitosphenoid", whereas other terms comprise general language words which acquire a new domain-specific meaning in a SL. Consequently, this author highlights that, despite the fact that SLs are derived from the general language (GL), they are lexically and semantically different. "The essential aim of the terminological lexicon is not the language itself" (Guilbert in Cabré & Sager, 1999, p. 11). Terminology is tightly linked to special languages and communication and addresses a variety of purposes that have to do with communication and information (Cabré, 1998, p. 11).

3. Research goals

The goal of this research is trying to determine the evolution of the lexical Astronomy-domain specificity in the CETA. That is, how much Astronomy-like lexical features occur in the English astronomic texts gathered in the CETA.

This might shed some light on: (i) the introduction rate of new astronomic specific vocabulary along time, (ii) lexical richness in English astronomic texts, (iii) the rate of new astronomic specific vocabulary along time, (iv) the potential lexical specific features of English astronomic texts, and (v) lexico-semantic text difficulty of English astronomic texts.

4. Methodology

With the aim of obtaining the Astronomy-specific vocabulary of the CETA, we extracted one-word terms (1-WTs) and multi-word terms (MWTs), based on two different approaches.

For single word-terms, we used the keyword methodology inbuilt in the *WordSmith* suit (V. 5.0; Scott, 2008). Keywords provide a useful way to characterize a text or a genre. Potential applications include: language teaching, forensic linguistics, stylistics, content analysis, text retrieval. The purpose of the *Keyword* tool is to locate and identify keywords in a given text. The keywords are worked out by first making a wordlist for a text, and a wordlist for a 'reference' corpus, then comparing the frequency of each word in the two lists. A reference corpus is any corpus chosen as a standard of comparison with your corpus. The reference corpus usually has to be quite large and of a suitable type for keywords to work. To do so, it compares the words in the text with a reference set of words usually taken from a large corpus of text. Any word which is found to be outstanding in its frequency in the text is considered 'key'. The keywords are presented in order of 'outstandingness'. If the word occurs say, 7% of the time in the small wordlist (text) and 8% of the time in the reference corpus, it will not turn out to be 'key', but if the word scores are 35% (in the text) and 6% (in the reference corpus), then it would be extremely 'key'.

For MWTs, we used *TerMine*[5] (Frantzi *et al.*, 2000). Many techniques for multi-word automatic term recognition (ATR) move from using only linguistic information (Ananiadou, 1988, 1994; Bourigaul, 1992) to incorporating statistical as well. Dagan and Church (1994), Daille *et al.* (1994), and Justeson & Katz (1995), Enguehard & Pantera (1994) use frequency of occurrence. Daille *et al.* (1994) and Lauriston (1996), propose the likelihood ratio for terms consisting of two words. *TerMine* uses a domain-independent method for multi-word ATR which aims to improve the extraction of nested terms. The method takes as input a corpus and produces a list of candidate MWTs. These are ordered by their termhood, also known as *C-value*. The *C-value* approach combines linguistic and

5 *TerMine* is freely available at <http://www.nactem.ac.uk/software/termine>.

statistical information. The linguistic information consists of the part-of-speech tagging of the corpus, the linguistic filter constraining the type of terms extracted, and the stoplist. The statistical part combines statistical features of the candidate string, in a form of measure that is also called *C-value*.

Lexical richness was obtained by means of the standardized type-token ratio (STTR) (Tweedie & Baayen, 1998; Scott, 2010).

The potential lexical specific features of English astronomic texts were obtained using three variables: (i) mean word length (Nam *et al.*, 2004); (ii) long words (>10 characters; Biber & Jones, 2005); and (iii) hapax legomena (Oakes, 2009).

Finally, for the lexico-semantic text difficulty, two parameters were used: (i) mean sentence length (Kelih *et al.*, 2006); and (ii) the automated readability index (Bruce & Rubin, 1988).

5. Data

In order to carry out the 1-WT extraction, we compared the CETA with a non-astronomic diachronically equivalent 'reference corpus'. The contemporary equivalent reference chosen was the Corpus of Late Modern English Texts, Extended Version (CLMETEV), which incorporates the full Corpus of Late Modern English Texts (CLMET) expanded to include an extra 5 million words of text, drawn from the Project Gutenberg, the Oxford Text Archive, or the Victorian Women Writers project.[6] In order to get a more fine tuned analysis of the keyword analysis, we fractioned the CETA and the CLMETEV into three chronologically equivalent and comparable segments. The time spans and corpus-chunks are given below (Table 1).

[6] More information on this corpus can be found in Vázquez *et al.* (forthcoming).

	Corpora	
Time span	**CETA**	**CLMETEV**
1710-1779	CETA-1	CLMET-1
1780-1849	CETA-2	CLMET-2
1850-1920	CETA-3	CLMET-3

Table 1. Corpus segments: CETA vs CLMETEV

6. Results

The preliminary sub-corpora data (tokens, types and standardized type-token ratios;[7] Table 2) show that overall the CETA corpus is lexically poorer, with STTRs ranging from 29.71 to 36.40. That is, usually 30 to 37 different word forms (types) are introduced per 1,000 words in late modern English Astronomy texts, compared to British-writer English, which is lexically more prolific (STTR 41.22-43.71).

	CETA-1	CETA-2	CETA-3	CLMETEV-1	CLMETEV-2	CLMETEV-3
Tokens	157,126	141,473	101,403	3,037,607	5,723,988	6,251,564
Types	7,355	7,200	7,109	48,833	52,611	66,485
STTR	29.71	33.75	36.40	41.22	43.71	43.41

Table 2. Comparing CETA and CLMETEV segment sizes

[7] The standardized type/token ratios (STTRs) have been calculated for the first 1,000 running words, then calculated afresh for the next 1,000, and so on to the end of the sub-corpus. A running average is computed, which means that STTR is an average type/token ratio based on consecutive 1,000-word chunks of text.

A more refined comparison contrasting the Astronomy-specific text periods with their chronologically equivalent reference sections evinces that lexical density is always lower in the Astronomy sections.

It is worth mentioning that whereas the STTRs in the reference sections are very stable across time, in the Astronomy-specific sections a constant increasing trend on lexical density can be appreciated. The CETA gets lexically richer across time: CETA-1 (1710-1779) accounts roughly for 30 new words per 1,000 tokens; CETA-2 (1780-1849) has on average 34 new types per 1,000 tokens; and CETA-2 (1850-1920) totals nearly 37 types for each 1,000 words. Astronomy texts seem to become lexically richer and also, likely, more complex and elaborated, probably due to the scientific advances in the field and/or the specialization of the audience these works were addressed to. A linear regression (Figure 2) on time and lexical density is statistically significant [$F(1,3) = 0.993$, $p = 0.038$]. This does not hold for the reference corpus (CLMETEV) where lexical density ranges from 42-44 types across 1710-1920; linear regression on time and lexical density is not significant [$F(1,3) = 0.806$, $p = 0.202$].

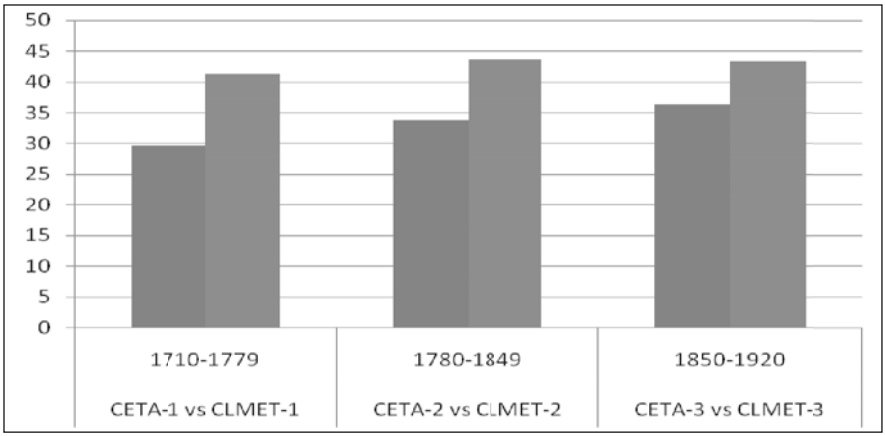

Figure 1. STTR Comparisons of the CETA and the CLMET

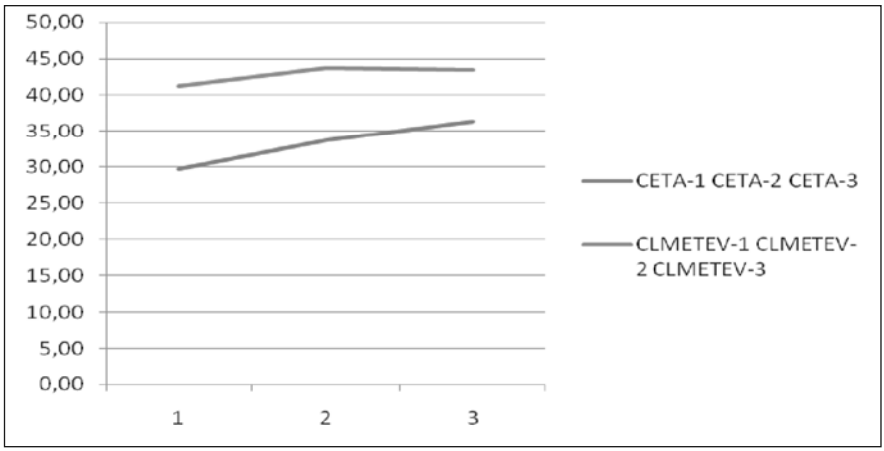

Figure 2. Linear regression: Lexical richness across time

To perform a similar analysis on the terms entailed in each of the three periods of the CETA, we extracted first 1-WTs and MWTs.[8] The overall amount of Astronomy terms is given below in Table 3.

	1-WTs	MWTs
CETA-1	930	250
CETA-2	830	193
CETA-3	813	199

Table 3. Astronomy keywords[9]

[8] Low frequency keywords (freq. < 3; *hapax legomena* and *hapax dislegomena*) were not considered.

[9] *Appendices A, B* and *C* account for the 100 most relevant one and multi-word terms.

A first look at the data might contradict the regression shown above. That is, while type-growth increases across time in the CETA, term-growth seems to decrease. In other words, astronomists seem to have started using less and less specific words in their writing across time; former Astronomy texts are more prolific in terminology than those written in later periods. This is not very plausible and, furthermore, contradicts common sense as sciences evolve and progress across time, and scientific findings are evidenced each time in more sophisticated writings, full of specific terminology unlikely to become accessible for laymen and/or non-experts in the field.

The problem with the above data is that it cannot be taken as it stands to make any straightforward comparisons; for one simple reason: the data has been extracted from different sub-parts of the CETA (CETA-1, CETA-2 and CETA-3), having each part a different size (see Table 2). We first need to normalize the data in order to contrast it accurately; as we did with the data in Table 2, using the sd-TTR, a standardized measure on type-growth.

Trying to normalize the data by means of relative frequencies or percentages would also be misleading as these measures depend on the size of the language sample used for measurement. Relative frequencies vary widely in accordance with the length of the text –or corpus of texts– which is being studied. For example, a 10,000 word article might have a TTR of 45; a shorter one might reach 80; 5 million words will probably give a ratio of about 1.7, and so on. Larger samples give always lower values. Such information is rather misleading and meaningless in most cases. The conventional ratios (*i.e.* relative frequencies) are informative, of course, if we are dealing with a corpus comprising equal-sized text segments (*e.g.* the LOB and Brown corpora). But in the real world, especially if the research focus is the text as opposed to the language, we shall probably be dealing with texts of different lengths and the conventional ratios will not help much. In a pilot project funded by the University of Reading, Richards and Malvern (Malvern & Richards, 1997; Richards & Malvern, 1996) found that this problem has distorted many research findings.

In order to overcome this problem, we shall be using a mathematical modelling approach for predicting the probability of new vocabulary being introduced (Sánchez & Cantos, 1997; 1998). This might give us a more reliable picture of how diverse the use of vocabulary is in a text or corpus. This measure

is not dependent on the total number of words in the sample. Succinctly, Sánchez & Cantos (1997, 1998) propose a quantitative measure that models the non-linear increase of types and lemmas, and by extension also other lexical items, such as terms. The original formula states that:

$$Types = K\sqrt{Tokens}$$

where *K* is a constant value specific to any text/corpus sample, etc. The modified formula for term density would thus be:

$$Terms = K_{Term}\sqrt{Tokens}$$

The index that reveals which text/corpus has a higher density in terminology is precisely the K_{Term}. So for CETA-1 we get:

$$Terms = K_{Term}\sqrt{Tokens}$$

$$1{,}180 = K_{Term}\sqrt{157{,}126}$$

$$K_{Term} = \frac{1{,}180}{\sqrt{157{,}126}} = \frac{1{,}180}{396.3912} = 2.9768$$

Table 4 and Figure 3 show the terminology density of all three periods of the CETA. Note that the K_{Term} has been calculated considering all terms (1-WTs + MWTs).

	Tokens	1-WT	MWT	Terms	K_{Term}
CETA-1	157,126	930	250	1180	2.9769
CETA-2	141,473	830	193	1023	2.7198
CETA-3	101,403	813	199	1012	3.1780

Table 4. Terminology density

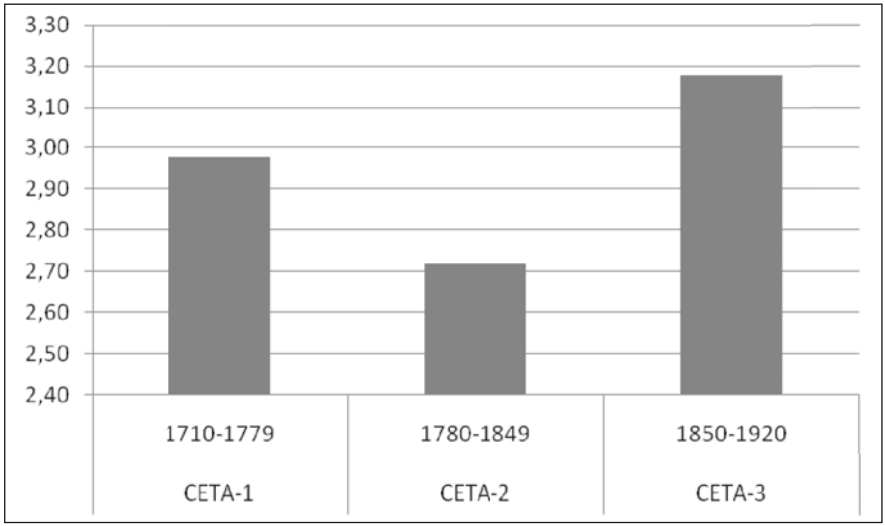

Figure 3. K_{Term}: Terminology density

The standardized data reveals that term-growth is not constant across time; Astronomy specific terms are more frequent in Astronomy books published between 1710 and 1779, than in the later period 1780-1850. A linear regression on year of publication and terminology density is not significant [$F(1,3) = 0.438$, $p = 0.356$]. This contrasts with the regression evidenced in type-growth. A partial explanation to this evidence might be the fact that astronomists in 1780-1849 resorted to describe their findings in their writings using probably more words related to general English and/or other scientific fields, not only that of Astronomy.

A detailed analysis, taking apart 1-WTs, MWTs and Ts, provides further evidence on this finding. Table 5 gives the *K-values* for 1-WTs, MWTs and overall Ts.

	$K_{1\text{-}WT}$	K_{MWT}	K_{Term}
CETA-1	2.3462	0.6307	2.9769
CETA-2	2.2067	0.5131	2.7198
CETA-3	2.5531	0.6249	3.1780

Table 5. Terminology densities relative to $K_{1\text{-}WT}$, K_{MWT} and K_{Term}

Surprisingly, CETA-2 (1780-1849) is overall the least prolific period regarding the usage on Astronomy terms, in general; this applies equally to the use of single-word terms (1-WTs) and multi-word terms (MWTs). This contrasts with the findings on type-growth, where CETA-2 is lexically denser than its former period (CETA-1). This might indicate, as stated above, that astronomists between 1780 and 1850 were more prone to use general English words and/or non-specific Astronomy terms in their scientific writings, than their former colleagues (1710-1779).

A *z-score* transformation[10] of all the *K-values* is even more conclusive and revealing (Table 6 and Figure 4). Note that the regression on time period and lexical density is only relevant regarding types [$F(1,3) = 0.993$, $p = 0.038$], but does not hold with Ts [$F(1,3) = 0.438$, $p = 0.356$], 1-WTs $F(1,3) = 0.594$, $p = 0.298$] or MWTs $F(1,3) = -0.044$, $p = 0.486$].

CETA-1 is more prolific in the usage of 1-WTs, MWTs and Ts in general, than the later periods CETA-2. The texts in the CETA-2 are less abundant in Astronomy-specific vocabulary than former texts on this scientific field (CETA-1) and later CETA-3 period, though it is more copious in incorporating new non-Astronomy specific words (types) in its repertoire than texts of the CETA-1 period. Similar to CETA-1, CETA-3 texts also conform to the general

[10] For a detailed information on how to calculate *z-scores* and their utility, see Woods *et al.* (1986: 184-7), Urdan (2005: 33-42), Gravetter and Wallnau (2007: 138-151), among others.

expected behaviour in that these writings are the latest and the most profuse ones, not only, in the usage of Astronomy-specific vocabulary (1-WTs and Ts in general), but also in the introduction of other new words in discourse (types).

Another interesting finding is that CETA-1 and CETA-3 used a similar amount of MWTs in their writings.

	Types	Terms	1-Word Terms	Multi-Word Terms
CETA-1	-1,0597	0,0825	-0,0793	0,6318
CETA-2	0,1395	-1,0402	-0,8810	-1,1500
CETA-3	0,9261	0,9607	1,1098	0,5439

Table 6. Standardized K-values

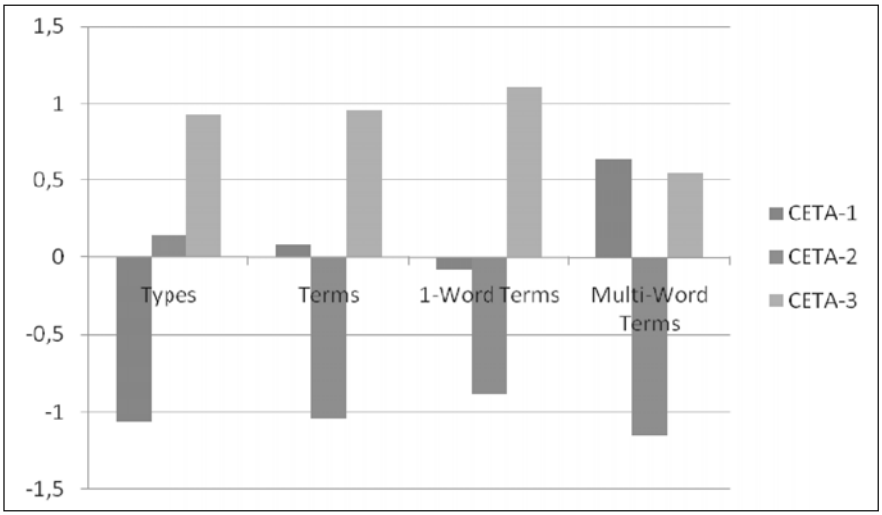

Figure 4. Standardized K-values

In addition, the mathematical model of Sánchez & Cantos (1997, 1998) has also a further advantage: it allows the potential prediction on term occurrence in larger texts chunks. We could, for instance, predict the number of terms in the three LME sub-corpora of 200,000 words, 1,000,000 words, and so on. Table 6 and Figure 5 show the prediction of term density in CETA-1, CETA-2 and CETA-3 of sizes varying from 100,000 to 1,000,000 tokens.

Tokens\Sub-corpus	CETA-1	CETA-2	CETA-3
100,000	907	852	988
200,000	1283	1205	1397
300,000	1571	1476	1711
400,000	1814	1704	1975
500,000	2028	1905	2209
600,000	2222	2087	2419
700,000	2400	2254	2613
800,000	2565	2410	2794
900,000	2721	2556	2963
1,000,000	2868	2694	3123

Table 6. Projection of term occurrence in the CETA

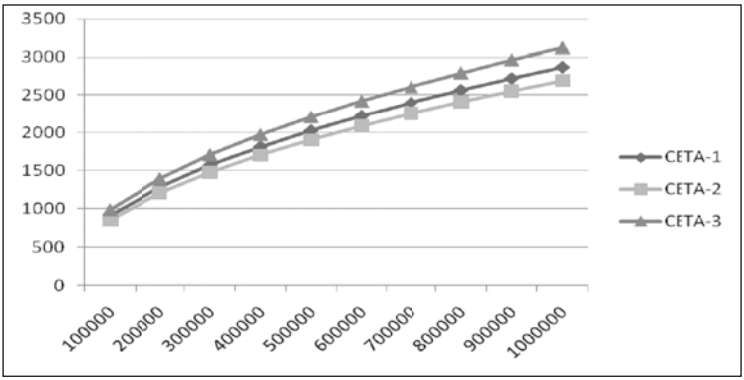

Figure 5. Projection of term occurrence in the CETA

Once again, CETA-2 exhibits an atypical behaviour as its subject specific vocabulary in Astronomy is clearly lower than would be expected compared to the former period: CETA-1. Consequently, CETA-2 is also the period which is the most prone to lexical closure, regarding Astronomy term usage.

In order to obtain potential lexical features specific to English astronomic texts, we used the following three measures: (i) mean word length (Nam *et al.*, 2004); (ii) long words (>10 characters; Biber & Jones, 2005); and (iii) hapax legomena (Oaks, 2009).

	Mean word length	*Long words*[11]	*Hapax legomena*[12]
CETA-1	4.2513	1.8428	1.8824
CETA-2	4.4569	2.6439	2.0118
CETA-3	4.5471	2.9015	2.8463

Table 7. Astronomy specific lexical features

11 Long words have been normalized to percentages.

12 Hapax legomena have been normalized to percentages.

This data reveals some interesting findings: (i) words tend to become longer across time, and (ii) there is an increase in the usage of rare words (hapax legomena) across time, too. However, the regressions on time period and the three lexical features specific to English astronomic texts are statistically not relevant regarding types: (i) mean word length [$F(1,3) = 0.975$, $p = 0.139$], (ii) long words [$F(1,3) = 0.959$, $p = 0.183$], and (iii) hapax legomena [$F(1,3) = 0.921$, $p = 0.254$]. Nevertheless, all three measures exhibit increases across time.

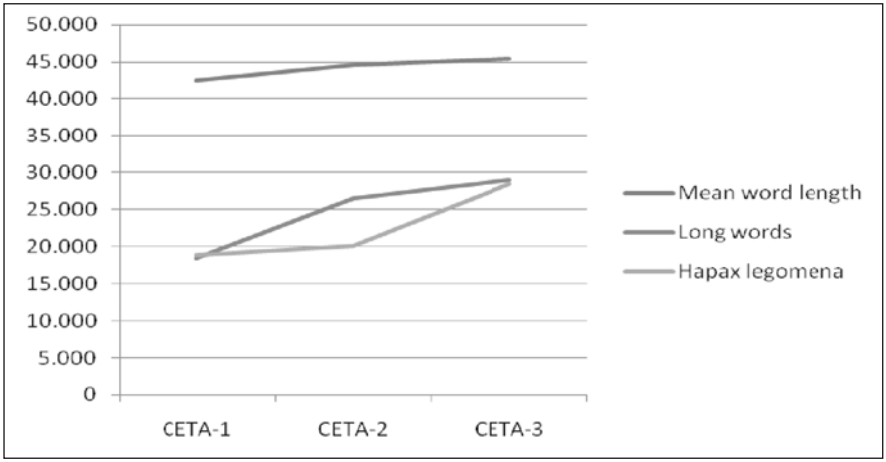

Figure 6. Increase of mean word length, usage of long words and hapax legomena in the CETA

Finally, the analysis of reading ease of Late Modern English astronomic texts, by means of the two parameters: (i) mean sentence length (Kelih *et al.*, 2006); and (ii) the automated readability index[13] (Bruce & Rubin, 1988), reveals that texts are

[13] The Automated Readability Index (ARI) is a readability test designed to gauge the understandability of a text. The formula for calculating the Automated Readability Index is:

$$4.71\left(\frac{\text{characters}}{\text{words}}\right) + 0.5\left(\frac{\text{words}}{\text{sentences}}\right) - 21.43$$

composed, on average, by shorter sentences across time and, in addition, these text become apparently also easier to read across time. However, neither the regression on time period and mean sentence length [$F(1,3) = -0.939$, $p = 0.224$], nor the regression on time period and reading ease [$F(1,3) = -0.868$, $p = 0.313$] are statistically relevant.

	Mean sentence length	*Automated Readability Index*
CETA-1	36,51	16,8486
CETA-2	34,53	16,8269
CETA-3	25,63	12,8018

Table 8. Lexico-semantic text difficulty

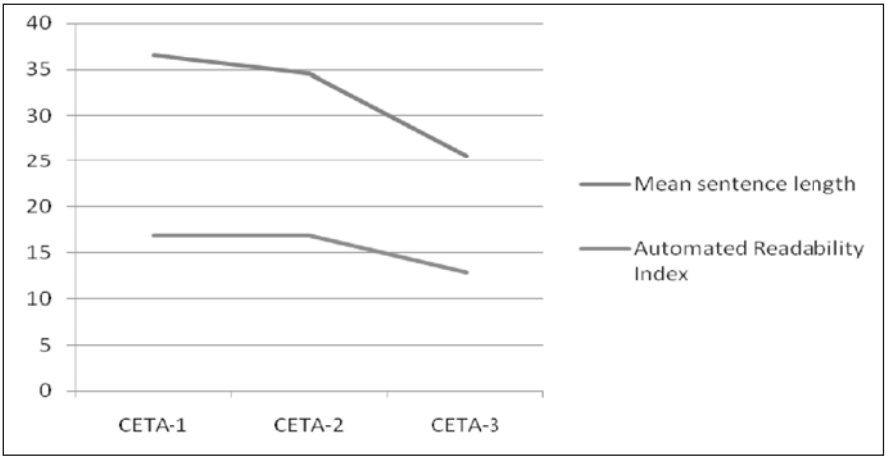

Figure 7. Decrease of mean sentence length and ARI in the CETA

7. Some final remarks

An interesting fact is that Astronomy specific vocabulary is not stable along the LME period. We have found important fluctuations along the three periods analyzed: 1710-1779, 1780-1849 and 1850-1920.

Astronomy specificity is particularly notorious in the last LME period, that is in texts ranging from 1850-1920, mainly regarding overall terms and one-word terms. However, texts dating between 1710 and 1779 are more prominent in multi-word terms.

The 'in-between' LME period (1780-1849) happens to be the least Astronomy -specific one.

A comparison of the 20-most prominent keywords across the three periods of the CETA shows:

- Total similarities (sun, earth, orbit, planets, motion, distance, planet and ecliptic),
- Partial similarities:
  - CETA-1 and CETA-2 (moon and horizon).
  - CETA-1 and CETA-3 (stars, Meridian, circle and Equator).
  - CETA-2 and CETA-3 (axis and surface).
- No similarities at all:
  - CETA-1 (globe, star, rectangle, circles, seconds and square).
  - CETA-2 (angle, plane, diameter, centre, bodies, line, comet and Venus).
  - CETA-3 (solar, endnote, telescope, plane, celestial, latitude).

	CETA-1	**CETA-2**	**CETA-3**
SUN	X	X	X
EARTH	X	X	X
ORBIT	X	X	X
MOTION	X	X	X
DISTANCE	X	X	X
PLANET	X	X	X
ECLIPTIC	X	X	X
MOON	X	X	
HORIZON	X	X	
STAR	X		X
MERIDIAN	X		X
CIRCLE	X		X
EQUATOR	X		X
AXIS		X	X
PLANE		X	X
SURFACE		X	X
GLOBE	X		
RECTANGLE	X		
SECOND	X		
SQUARE	X		
ANGLE		X	
DIAMETER		X	

CENTRE		X	
BODY		X	
LINE		X	
COMET		X	
VENUS		X	
SOLAR			X
ENDNOTE			X
TELESCOPE			X
CELESTIAL			X
LATITUDE			X

Table 9. Most prominent Astronomy Keywords across Late Modern English

Differences in the usage of Astronomy lexicon across LME might be due to new discoveries and/or research interest in different time periods.

A more detailed consistency analysis of the usage of Astronomy keywords across LME periods shows that there are more similarities between CETA-1 and CETA-2 (9.56%) than between CETA-1 and CETA-3 (2.61%); and also CETA-2 is closer to CETA-3 (7.84%). This clearly attests that advances in Astronomy during 1850-1920 evidence little similarities with the former period 1710-1780. There is a clear transition regarding Astronomy specific use of the vocabulary along the LME period. This fact becomes also relevant if we focus our attention on the total number of keywords exclusively used in each period. Undoubtedly, the period between 1710 and 1780 is most prolific in the use of Astronomy terminology (28.02%), followed by the 1850-1920 period (21.07). The in-between period (1780-1850) 'borrows' most of the items from 1710-1750 and is also less fertile in forming new words related to Astronomy. This might also be a hint pointing towards a less fruitful period in Astronomy research.

	Amount of KWs	**% of KWs**
CETA-1, CETA-2, CETA-3	209	15.62
CETA-1, CETA-2	128	9.56
CETA-1, CETA-3	35	2.61
CETA-2, CETA-3	105	7.84
CETA-1	375	28.02
CETA-2	204	15.24
CETA-3	282	21.07

Table 10. Consistency analysis of LME Astronomy keyword usage

An interesting paradox regarding the lexical repertoire of LME astronomic texts is that while their authors increased the usage of longer and rarer words (mean word length, long words and hapax legomena) in their research findings, the reading difficulty of their writings decrease (see sentence length and ARI). There seems to be an inverse relationship between long/rare words and reading difficulty, which is an interesting paradox, worth looking into.

Bibliographical References

Alonso-Almeida, F. & **Sánchez-Cuervo, M.** (2009). The vernacularisation of Medieval medical texts. In Bravo, S. *et al. Estudios de traducción* (pp. 191-207). Frankfurt am Main: Peter Lang.

Ananiadou, S. (1988). *Towards a methodology for automatic term recognition*. PhD thesis, University of Manchester Institute of Science and Technology.

Ananiadou, S. (1994). A methodology for automatic term recognition. *Proceedings of the 15th International conference on computational linguistics, COLING,* 94, 1034-1038.

Atkinson, D. (1992). The evolution of medical research writing from 1735 to 1985: The case of the Edinburgh medical journal. *Applied Linguistics, 113*, 337-374.

Atkinson, D. (1996). The Philosophical Transactions of the Royal Society of London, 1675-1975: A sociohistorical discourse analysis. *Language and Society, 5*, 333-371.

Biber, D. & **Jones, J. K.** (2005). Merging corpus linguistic and discourse analytic research goals: Discourse units in biology research articles. *Corpus Linguistics and Linguistic Theory, 1*, 151-182.

Bourigaul, D. (1992). Surface grammatical analysis for the extraction of terminological noun phrase. *Proceedings of the fifteenth international conference on computational linguistics.*

Brown, G. & **Yule , G.** (1983). *Discourse analysis.* Cambridge: CUP.

Bruce, B. & **Rubin, A.** (1988). Readability formulas: Matching tool and task. In A. Davison & G. M. Green (Eds.), *Linguistic complexity and text comprehension: Readability issues reconsidered* (pp. 5-22). Hillsdale, New Jersey: Erlbaum.

Cabré, M. T. (1993). *La terminología: Teoría, metodología, aplicaciones.* Barcelona: Ed. Antártida.

Cabré, M. T. (1998). *Terminology. Theory, methods and applications.* Amsterdam: John Benjamins.

Cabré, M. T. & **Sager, J. C.** (1999). *Terminology. Theory, methods and applications.* Amsterdam: John Benjamins.

Dagan, I. & **Church, K.** (1994). Termight: Coordinating humans and machines in bilingual terminology acquisition. *Machine Translation*, vol. 12 (1/2), 89-107.

Dagan, I. & **Church, K.** (1995). Termight: Identifying and translating technical terminology. *Proceedings of the 4th conference on applied natural language processing*, 34-40.

Daille, B., Gaussier, E. & **Langé, JM.** (1994). Towards automatic extraction of monolingual and bilingual terminology. In *Proceedings of COLING, 94,* 515-521.

Denison, D. (1998). Syntax. In Romaine, S. (Ed.) *The Cambridge history of the English language IV: 1776-1997* (pp. 92-329). Cambridge: Cambridge University Press.

Enguehard, C. & **Pantera, L.** (1994). Automatic natural acquisition of a terminology. *Journal of Quantitative Linguistics, 2*(1), 27-32.

Frantzi, K., Ananiadou, S. & **Mima, H.** (2000). Automatic recognition of multi-word terms: the C-value/NC-value method. *International Journal on Digital Libraries, 3*, 115-130.

Gravetter, F. & **Wallnau, L.** (2007). *Essentials of statistics for the behavioral science*. Belmont, CA: Thomson Higer Education.

Justeson, J. S. & **Katz, S. M.** (1995). Technical terminology: Some linguistic properties and an algorithm for identification in text. *Natural Language Engineering, 1,* 9-27.

Kelih, E., Grzybek, P., Antic, G. & **Stadlober, E.** (2006). Quantitative text typology: The impact of sentence length. In M. Spiliopoulou *et al.* (Eds.) *From data and information analysis to knowledge engineering* (pp. 382–389). Berlin: Springer Verlag.

Lareo-Martín, I. & **Montoya-Reyes, A.** (2007). Scientific writing: Following Robert Boyle's principles in experimental essays, 1704 & 1998. *Revista Alicantina de Estudios Ingleses, 20*, 119-137.

Lauriston, A. (1996). *Automatic term recognition: Performance of linguistic and statistical learning techniques* (PhD thesis). Manchester: University of Manchester Institute of Science and Technology.

Malvern, D. D. & **Richards, B. J.** (1997). *Quantifying lexical diversity in the study of language development*. Reading: University of Reading, The New Bulmershe Papers.

Milroy, J. (1992). *Linguistic variation & change*. Oxford: Basil Blackwell.

Moskowich-Spiegel Fandiño, I. (Forthcoming) "A smooth homogeneous globe" in CETA: Compiling Late Modern Astronomy texts in English. In Vázquez, N. (Ed.) *Creation and use of historical English corpora in Spain*. Newcastle: Cambridge Scholars.

Moskowich-Spiegel Fandiño, I. & **Crespo-García, B.** (2007). Presenting the Coruña Corpus: A collection of samples for the historical study of English scientific writing. In Pérez Guerra, J. *et al.* (Eds.) *Of varying language and opposing creed: New insights into Late Modern English* (pp. 341-357). Bern: Peter Lang.

Moskowich-Spiegel Fandiño, I. & **Crespo-García, B.** (2009). The limits of my language are the limits of my world: The scientific lexicon from 1350 to 1640. *SKASE Journal of Theoretical Linguistics, 6*(1), 45-58.

Nam, Y. H., Park, S. H., Ha, T. K. & **Jeon, Y. H.** (2004). Preprocessing of digital audio data for mobile audio codecs. [<http://www.freepatentsonline.com/y2004/0128126.html>]

Nevalainen, T. (1999). Early Modern English lexis and semantics. In R. Lass, *The Cambridge history of the English language* Vol. 3 (pp. 1476-1776). Cambridge: CUP.

Norri, J. (1992). *Names of sicknesses in English, 1400-1550: An exploration of the lexical field. Annales academiae scientiarum fennicae dissertationes humanarum litterarum* 63. Helsinki.

Norri, J. (1998). *Names of body parts in English, 1400-1550. Annales academiae scientiarum fennicae dissertatione. Humaniora* 291. Helsinki.

Oakes, M. P. (2009). Corpus linguistics and stylometry. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: An international handbook* (pp. 1070-1090). Berlin: Mouton de Gruyter.

Richards, B. J. & **Malvern, D. D.** (1996). Swedish verb morphology in language impaired children: Interpreting the type-token ratios. *Logopedics Phoniatrics Vocology*, *21*(2), 109-111.

Romaine, S. (1994). *Language in society: An introduction to sociolinguistics*. Oxford: OUP.

Rydén, M. (1984). The study of eighteenth century English syntax. In J. Fisiak (Ed.) *Historical syntax* (pp. 509-520). Berlin: Mouton Publishers.

Sager, J. C. (1990). *A practical course in terminology processing*. Amsterdam: John Benjamins Publishing Company.

Sánchez, A. & **Cantos, P.** (1997). Predictability and representativeness of words, word forms and lemmas in linguistic corpora. A case study based on the analysis of the *CUMBRE corpus*: An 8-million word corpus of contemporary Spanish". *International Journal of Corpus Linguistics, 2*(2), 259-280.

Sánchez, A. & **Cantos, P.** (1998). El ritmo incremental de palabras nuevas en los repertorios de textos. Estudio experimental comparativo basado en dos corpus lingüísticos equivalentes de cuatro millones de palabras de las lenguas

inglesa y española y en cinco autores de ambas lenguas. *ATLANTIS, 19*(2), 205-223.

Scott, M. (2008). *Wordsmith tools version 5*. Liverpool: Lexical Analysis Software.

Scott, M. (2010). *WordSmith tools.* [<http://www.lexically.net/downloads/version5/WordSmith.pdf>]

Spasic, I., Ananiadou, S., McNaught, J. & **Kumar, A.** (2005). Text mining and ontologies in biomedicine: Making sense of raw text. *Briefings in Bioinformatics, 6*(3), 239-251.

Taavitsainen, I. (2001). Language history and the scientific register. In H. J. Diller & M. Görlach (Eds.), *Towards a history of English as a history of genres* (pp. 185-202). Heidelberg: Winter.

Taavitsainen, I., Pahta, P., Leskinen, N., Ratia, M. & **Suhr, C.** (2002). Analysing scientific thought-styles: What can linguistic research reveal about the History of Science? In H. Raumolin-Brunberg, M. Nevala, A. Nurmi & M. Rissanen (Eds.), *Variation past and present, VARIENG Studies on English for Terttu Nevalainen* (pp. 251-270). Helsinki: Société Néophilologique.

Tweedie, F. & **Baayen, H.** (1998). How variable may a constant be? Measures of lexical richness in perspective. *Computers in the Humanities, 32,* 323-352.

Urda, T. (2005). *Statistics in plain English*. Mahwah, New Jersey: Lawrence Erlbaum.

Vázquez, N. *et al.* (Forthcoming). A descriptive approach to English historical corpora in the 21st Century. *IJES, 11*(2). *New Developments in Corpus Linguistics.*

Woods, A., Fletcher, P., Hughes, A., Austin, P., Bresnan, J., Comrie, B., Crain, S., Dressler, W., Ewen, C. & **Lass, R.** (1986). *Statistics in language studies.* Cambridge: CUP.

Appendix A: Terms in CETA-1

N	1-WTs	Keyness	N	1-WTs	Keyness
1	SUN	6070.6	51	DECLINATION	592.7
2	EARTH	5505.2	52	HOURS	582.9
3	MOON	3512.7	53	PARALLEL	582.8
4	STARS	2821.9	54	SURFACE	573.5
5	HORIZON	2301.6	55	PARALLAX	571.1
6	MERIDIAN	1966.7	56	CONSTELLATION	563.5
7	CIRCLE	1946.5	57	DEGREES	559.6
8	GLOBE	1640.1	58	RATIO	550.3
9	ORBIT	1496.1	59	DIAMETER	537.6
10	STAR	1414.7	60	NODES	512.3
11	RECTANGLE	1413.3	61	AB	494.1
12	EQUATOR	1365.1	62	VERTICAL	476.9
13	CIRCLES	1324.2	63	MINUTES	461.6
14	PLANETS	1221.0	64	SATURN	459.2
15	SECONDS	1206.3	65	LATITUDE	458.9
16	MOTION	1170.2	66	NODE	447.5
17	SQUARE	1167.0	67	DISTANCES	432.0
18	DISTANCE	1137.5	68	ENDNOTE	429.8
19	PLANET	1127.6	69	CONSTELLATIONS	429.1
20	CALLED	1117.0	70	CANCER	418.0
21	ECLIPTIC	1079.2	71	HEMISPHERE	417.4
22	CENTER	1077.5	72	ZENITH	416.5
23	ANGLE	1067.9	73	ABCD	412.2

24	AXIS	918.4	74	PLACES	404.3
25	FIG	900.6	75	ASTRONOMERS	401.2
26	JUPITER	898.4	76	CELESTIAL	390.5
27	SUN'S	892.8	77	MARS	382.8
28	POLES	891.4	78	SECTOR	378.3
29	SIGNS	890.3	79	DIURNAL	376.0
30	POLE	864.0	80	SPACE	371.6
31	HEAVENS	852.4	81	ECLIPSES	364.1
32	TA	829.1	82	YEAR	360.5
33	LINE	817.3	83	DAYS	359.1
34	ROUND	795.5	84	HOUR	356.4
35	EQUAL	777.4	85	ELLIPSE	354.9
36	POINT	767.3	86	ATMOSPHERE	351.5
37	EQUINOCTIAL	748.8	87	AF	347.4
38	MOON'S	734.1	88	MOVE	336.2
39	EARTH'S	718.3	89	BODIES	329.5
40	PLANE	712.7	90	DESCRIBING	329.2
41	ECLIPSE	711.6	91	GLOBES	319.8
42	ANGLES	660.8	92	ASCENSION	319.8
43	SPHERE	658.5	93	POLAR	318.0
44	FIXED	657.8	94	CONJUNCTION	317.1
45	POINTS	655.8	95	SIGN	313.6
46	ECLIPTICK	653.6	96	MAGNITUDE	310.6
47	SATELLITE	632.0	97	CONTAINED	307.3

48	PLACE	627.6	98	AE	306.3
49	SHADOW	618.6	99	FRAGMENTGREEK	306.2
50	ARIES	595.3	100	ARCH	305.4

N	**MWTs**	**C-Score**	**N**	**MWTs**	**C-Score**
1	FIXED STARS	58	51	LUNAR ECLIPSE	10
2	EQUAL PART	44	52	SPACE Q	9
3	CENTRAL FORCE	30	53	HOUR INDEX	9
4	DIURNAL MOTION	29.5	54	REVOLUTION ROUND	9
5	POLAR CIRCLE	28	55	STARRY HEAVEN	9
6	CENTRIPETAL FORCE	27	56	WINTER SOLSTICE	9
7	ANNUAL MOTION	27	57	PROPER PLACE	9
8	CELESTIAL GLOBE	26	58	CARDINAL POINT	9
9	BRIGHT STAR	26	59	POINT G	8
10	RADICAL PLACE	24	60	EAST SIDE	8
11	SOUTH POLE	24	61	RECTANGLE TMB	8
12	FIXED STAR	23	62	GEOCENTRIC MOTION	8
13	SUBDUPLICATE RATIO	23	63	ANGLE FEG	8
14	NORTH POLE	22	64	NORTH SIDE	8
15	DISTURBING FORCE	22	65	SOUTH SIDE	8
16	BRASS MERIDIAN	21	66	RECTANGLE AEB	8

17	APPARENT MOTION	20	67	NORTHERN HEMISPHERE	8
18	TERRESTRIAL GLOBE	20	68	HORIZONTAL LINE	8
19	VERTICAL CIRCLE	19.5	69	SMALL STAR	8
20	STRAIT LINE	19	70	TEMPERATE ZONE	8
21	PRIMARY PLANET	19	71	EQUINOCTIAL COLURE	8
22	FULL MOON	18	72	SOUTH DECLINATION	8
23	ANNUAL PARALLAX	18	73	APPARENT DISTANCE	8
24	APPARENT DIAMETER	18	74	VERNAL EQUINOX	8
25	RIGHT ASCENSION	18	75	AUTUMNAL EQUINOX	8
26	ELLIPSE ABCD	17	76	FOURTH SATELLITE	8
27	HEAVENLY BODY	16	77	NORTH LATITUDE	8
28	COMMON JULIAN YEAR	15.8	78	INFERIOR PLANET	8
29	RATIONAL HORIZON	15.5	79	SPACE AD	8
30	EQUINOCTIAL POINT	15	80	TORRID ZONE	8
31	NEW MOON	14	81	UNFORMED STAR	8
32	SUPERIOR PLANET	14	82	CENTRE T	8
33	SOUTHERN HEMISPHERE	12	83	SECONDARY PLANET	8

34	LONG DAY	12	84	CIRCLE ABL	7
35	ANNUAL ORBIT	12	85	SPHERICAL SURFACE	7
36	PARALLEL LINE	11	86	SUN S	7
38	SMALL CIRCLE	11	88	SOUTH LATITUDE	7
39	LEFT HAND	11	89	DOMINICAL LETTER	7
40	POLAR STAR	11	90	BRASS CIRCLE	7
41	APPARENT PLACE	11	91	EARTHLY GLOBE	7
42	OBLIQUE ASCENSION	11	92	POLE STAR	7
43	SOLSTITIAL COLURE	10	93	COPERNICAN SYSTEM	7
44	NAKED EYE	10	94	SUMMER SOLSTICE	7
45	SPACE DM	10	95	ANNUAL COURSE	7
46	BRAZEN MERIDIAN	10	96	CELESTIAL LIGHT	7
47	VISIBLE HORIZON	10	97	MILKY WAY	7
48	DIURNAL PARALLAX	10	98	OBLIQUE SPHERE	7
49	CENTRIFUGAL FORCE	10	99	WEST POINT	7
50	GREAT CIRCLE	10	100	WEST SIDE	7

Appendix B: Terms in CETA-2

N	1-WTs	Keyness	N	1-WTs	Keyness
1	SUN	4279.8	51	COMETS	412.5
2	EARTH	3391.6	52	STARS	409.8
3	MOON	3052.0	53	ALTITUDE	381.1
4	ORBIT	2488.4	54	FIG	377.3
5	DISTANCE	1588.0	55	ENDNOTE	373.7
6	ANGLE	1580.9	56	MILES	369.9
7	MOTION	1442.8	57	ROUND	369.0
8	PLANE	1396.4	58	DIFFERENCE	359.4
9	PLANETS	1298.3	59	COINCIDE	358.9
10	MOON'S	1219.3	60	SATURN	358.1
11	ECLIPTIC	1219.3	61	LIGHT	357.9
12	DIAMETER	1206.0	62	SOLAR	355.8
13	HORIZON	1134.3	63	ADJUSTMENT	353.6
14	PLANET	1055.6	64	ASTRONOMERS	349.9
15	CENTRE	1049.5	65	TELESCOPE	343.6
16	AXIS	1045.1	66	BODY	340.6
17	SURFACE	897.5	67	ANGLES	338.4
18	BODIES	850.3	68	RADIUS	335.7
19	EARTH'S	846.0	69	EQUATOR	332.1
20	LINE	799.8	70	POINT	319.5
21	COMET	790.4	71	MERCURY	317.7
22	SUN'S	770.8	72	COR	312.5
23	VENUS	750.6	73	OBSERVATION	311.8

24	DISTANCES	744.0	74	PARTS	304.0
25	QUADRANT	725.1	75	LUNAR	297.7
26	FORCE	702.7	76	POLE	293.7
27	INDEX	659.3	77	ATMOSPHERE	293.1
28	ORBITS	652.3	78	MOONS	287.4
29	MOTIONS	591.7	79	FORCES	283.6
30	PARALLEL	585.8	80	TIMES	279.7
31	JUPITER	583.5	81	ARIES	278.7
32	PERPENDICULAR	578.2	82	EPICYCLE	278.7
33	ARC	568.6	83	REVOLVE	277.4
34	EQUAL	568.1	84	CELESTIAL	276.1
35	PARALLAX	558.5	85	ECLIPSES	274.9
36	VELOCITY	509.3	86	VISIBLE	273.3
37	NODE	506.7	87	QUADRATURE	272.3
38	RATIO	501.8	88	AREAS	272.3
39	CENTER	487.7	89	MARS	271.4
40	NODES	487.7	90	MOVE	269.4
41	APPARENT	486.0	91	ASTRONOMY	266.7
42	LATITUDE	462.3	92	RETROGRADE	265.2
43	ECLIPSE	461.1	93	PROPORTIONAL	262.7
44	MERIDIAN	457.0	94	LONGITUDE	260.7
45	SEMI	447.9	95	STAR	260.7
46	GRAVITY	446.3	96	AD	257.8
47	REVOLUTION	442.6	97	CIRCLE	254.1

48	ANGULAR	431.5	98	APSIDES	253.3
49	AB	413.1	99	BD	253.3
50	DEGREES	4279.8	100	GLOBE	412.5

N	MWTs	C-Score	N	MWTs	C-Score
1	FULL MOON	34.7	51	PHYSICAL ASTRONOMY	8
2	APPARENT DIAMETER	30	52	LUMINOUS RING	8
3	CENTRIPETAL FORCE	27	53	LUNAR ORBIT	8
4	HORIZON GLASS	26	54	APPARENT SEMI-DIAMETER	8
5	SUPERIOR PLANET	23	55	AUTUMNAL FULL MOON	7.924.812
6	INDEX GLASS	23	56	DIURNAL REVOLUTION	7.5
7	ANGULAR DISTANCE	23	57	CHRISTIAN ERA	7
8	COMMON CENTRE	22	58	FORE HORIZON	7
9	HEAVENLY BODY	22	59	DARK SHADOW	7
10	INFERIOR PLANET	22	60	FORCE MK	7
11	ANGULAR VELOCITY	21	61	SMALL CIRCLE	7
12	APPARENT MOTION	19	62	TANGENT AD	7
13	DISTANT OBJECT	17	63	ANNUAL REVOLUTION	7

14	ULTIMATE RATIO	17	64	DIRECT MOTION	7
15	CELESTIAL BODY	17	65	PROPORTIONAL DISTANCE	7
16	HORIZONTAL PARALLAX	17	66	REAL SIZE	7
17	NAKED EYE	15	67	JULIAN CALENDAR	7
18	NORTH POLE	14	68	SOUTH POLE	7
19	FORE OBSERVATION	14	69	SIR ISAAC NEWTON	6.924.812
20	REAL MOTION	13	70	POINT B	6.5
21	RETROGRADE MOTION	13	71	POINT D	6
22	BACK OBSERVATION	12	72	POINT L	6
23	DIURNAL ROTATION	12	73	ANGULAR MOTION	6
24	SOLAR SYSTEM	12	74	LUNAR ATMOSPHERE	6
25	DUPLICATE RATIO	12	75	POLAR CIRCLE	6
26	APPARENT PLACE	12	76	JULIAN PERIOD	6
27	BODY T	12	77	PARALLEL LINE	6
28	DISTURBING FORCE	12	78	STRAIGHT LINE	6
29	SYNODIC REVOLUTION	11	79	STARRY SPHERE	6
30	ANNUAL MOTION	11	80	DIURNAL MOTION	6
31	CENTRIFUGAL FORCE	11	81	CENTRAL FORCE	6
32	DOMINICAL LETTER	11	82	COMMON CENTER	6

33	MERIDIAN ALTITUDE	11	83	APPARENT MAGNITUDE	6
34	PLANETARY ORBIT	10	84	MUTUAL ACTION	6
35	EQUAL PART	10	85	OBSCURE PART	6
36	IMMENSE DISTANCE	10	86	SINGLE FORCE	6
37	ANNUAL ORBIT	10	87	ZENITH DISTANCE	6
38	RADIUS VECTOR	10	88	REAL DISTANCE	6
39	DARK SPOT	9	89	LUNAR ECLIPSE	6
40	VISIBLE HORIZON	9	90	SOLAR ECLIPSE	6
41	GEORGIUM SIDUS	9	91	NAUTICAL ALMANAC	6
42	BACK HORIZON	9	92	UNIFORM MOTION	6
43	CELESTIAL OBJECT	9	93	EVENING STAR	6
44	CELESTIAL MOTION	9	94	ABLATITIOUS FORCE	6
45	ARC ACB	8.5	95	PLANETARY MOTION	6
46	PRIMARY PLANET	8	96	LINE AB	5
47	PERIODICAL REVOLUTION	8	97	HARVEST MOON	5
48	EQUAL AREA	8	98	PLANE PG	5
49	COPERNICAN SYSTEM	8	99	ENLIGHTENED SIDE	5
50	SMALL ANGLE	8	100	CENTRAL PART	5

Appendix C: Terms in CETA-3

N	1-WTs	Keyness	N	1-WTs	Keyness
1	SUN	1729.8	51	NEBULÆ	345.0
2	EARTH	1697.8	52	DIURNAL	336.9
3	SOLAR	1414.8	53	DENUDATION	336.8
4	ENDNOTE	1366.3	54	ASTRONOMY	327.7
5	TELESCOPE	1117.0	55	NEBULA	319.2
6	DISTANCE	1037.5	56	VERTICAL	314.6
7	ORBIT	1025.7	57	VENUS	306.5
8	STARS	1019.7	58	POLAR	299.2
9	MERIDIAN	963.9	59	SPECTRUM	297.3
10	AXIS	923.0	60	OCULAR	293.8
11	PLANE	912.8	61	ARC	292.4
12	EQUATOR	892.0	62	SYSTEM	291.5
13	MOTION	874.8	63	RAYS	288.7
14	EARTH'S	836.7	64	FIGURE	286.1
15	CELESTIAL	814.0	65	EQUATION	285.2
16	PLANET	801.7	66	ASTRONOMERS	281.4
17	CIRCLE	798.9	67	POINTS	273.6
18	PLANETS	736.0	68	DENSITY	272.3
19	SURFACE	727.5	69	COMPUTED	272.0
20	LATITUDE	724.3	70	ASTR	271.0
21	SUN'S	705.0	71	OBSERVER	266.0
22	ECLIPTIC	607.8	72	HEAT	265.3
23	ANGLE	593.2	73	EYE	257.9

24	NOTE	585.6	74	MOON'S	257.5
25	CANALS	585.1	75	OBJECT	257.3
26	HORIZON	585.0	76	DISTANCES	256.2
27	DIAMETER	554.6	77	VERNIER	254.6
28	STAR	531.8	78	MAGNIFYING	254.5
29	ROTATION	514.8	79	VOL	254.5
30	BODIES	491.4	80	PARALLEL	247.2
31	PARALLAX	476.4	81	APPARENT	243.7
32	LUNAR	472.2	82	FOCUS	241.3
33	POLE	457.2	83	NORTH	238.9
34	CENTRE	435.6	84	ORBITS	237.8
35	EQUINOCTIAL	409.0	85	MILES	234.0
36	ZENITH	407.4	86	DISC	231.5
37	GLOBE	400.3	87	MARS	228.3
38	SPOTS	387.0	88	ASTRONOMICAL	226.4
39	EQUAL	384.5	89	TERRESTRIAL	223.1
40	LENS	378.4	90	INCLINATION	221.3
41	VELOCITY	371.0	91	LIMB	212.0
42	OBSERVATIONS	367.9	92	THEORY	211.8
43	QUOTATION	365.4	93	SPACE	211.4
44	SPHERE	358.5	94	HEAVENS	209.2
45	LINES	357.9	95	LONGITUDE	209.2
46	MEASURED	357.5	96	TRANSIT	206.8
47	POINT	356.1	97	CENTRIFUGAL	206.0

48	MOON	353.9	98	PRECESSION	205.3
49	RADIUS	352.8	99	CIRCLES	202.9
50	MERCURY	345.2	100	VISIBLE	200.1

N	**MWTs**	**C-Score**	**N**	**MWTs**	**C-Score**
1	CELESTIAL SPHERE	52	51	MERIDIAN ZENITH DISTANCE	7.924.812
2	PROPER PLANE	49.5	52	LUNAR PROPER PLANE	7.924.812
3	SOLAR SYSTEM	45	53	CENTRAL LATITUDE	7
4	EYE PIECE	42.5	54	EQUATORIAL RADIUS	7
5	HEAVENLY BODY	31	55	DIURNAL PATH	7
6	CENTRIFUGAL FORCE	26	56	SOLAR SURFACE	7
7	ZENITH DISTANCE	25.5	57	SOLAR ATMOSPHERE	7
8	VERTICAL CIRCLE	22	58	POLE STAR	7
9	POLAR DISTANCE	21.5	59	SEDIMENTARY ROCK	7
10	FOCAL LENGTH	19	60	BRIGHT STAR	7
11	OBJECT GLASS	17	61	CIRCLE DIVISION	7
12	LUNAR ORBIT	17	62	APPARENT ANGULAR MAGNITUDE	633.985
13	INVARIABLE PLANE	16	63	RATIONAL HORIZON	6

14	HORIZONTAL PARALLAX	16	64	CONTRARY DIRECTION	6
15	NAKED EYE	15	65	INSTRUMENTAL AZIMUTH	6
16	SMALL CIRCLE	13	66	GRAVITATION THEORY	6
17	OPTICAL AXIS	13	67	CENTRAL ZENITH	6
18	CELESTIAL EQUATOR	13	68	NATURAL RELIGION	6
19	CIRCULAR ORBIT	13	69	SOUTH POINT	6
20	EQUATORIAL HORIZONTAL PARALLAX	12.6	70	DARK BODY	6
21	DIURNAL MOTION	12	71	AVERAGE DENSITY	6
22	CUBIC FOOT	12	72	SQUARE MILE	6
23	SMALL ANGLE	12	73	DARK REGION	6
24	STRATIFIED ROCK	12	74	CELESTIAL POLE	6
25	STRAIGHT LINE	11	75	MODERN SCIENCE	6
26	EQUAL PART	11	76	PROBABLE ERROR	6
27	SOLAR SPOT	11	77	SOLAR RADIATION	6
28	NORTH POLE	10	78	SMALL PORTION	6
29	PRECESSIONAL VELOCITY	10	79	APPARENT SEMI-DIAMETER	6
30	BLACK SPOT	10	80	MERE POINT	6
31	VISUAL RAY	10	81	NEW YORK	6
32	SOUTH POLE	10	82	TERRESTRIAL EQUATOR	6

33	RIAL DENUDATION	10	83	MERIDIAN ALTITUDE	6
34	SIR JOHN HERSCHEL	9.5	84	ENTIRE REVOLUTION	6
35	DARK LINE	9	85	GASEOUS CONDITION	6
36	SOLAR SPECTRUM	9	86	CENTRAL ZENITH DISTANCE	5.924.812
37	VERNAL EQUINOX	9	87	SMALL ILLUMINATED CIRCLE	533.985
38	SOLAR ECLIPSE	9	88	CELESTIAL SPHERE.	5
39	ANGULAR DISTANCE	9	89	PHYSICAL CONSTITUTION	5
40	NODAL VELOCITY	9	90	SPHERICAL SURFACE	5
41	MICROMETER HEAD	9	91	VISIBLE POLE	5
42	SOLAR PARALLAX	9	92	SQUARE ROOT	5
43	SECULAR CHANGE	8	93	BRIGHT LINE	5
44	COMPTES RENDUS	8	94	SPHERICAL ASTRONOMY	5
45	SOLAR NEBULA	8	95	AVERAGE RATE	5
46	SOLAR ORB	8	96	ANIMAL LIFE	5
47	CELESTIAL OBJECT	8	97	MIDDLE POINT	5
48	BASIC LINE	8	98	SYRTIS MAJOR	5
49	VERTICAL LINE	8	99	COMMON NODE	5
50	BASE LINE	8	100	REAL MOTION	5